Network Working Group
RFC: 316
NIC: 9346

D. B. McKay, IBM
A. P. Mullery, IBM
February 23 & 24, 1972

Categories:
Obsoletes:
Updates:

RCVD 3APR72

ARPA Network Data Management Working Group

The meeting had two different phases. The first included presentations of applications of networks and development work in the design to allow data sharing in a computer network, the second was a working meeting in which was discussed what the data management working group should do.

Phase I

John Senior, Univ. of Penn. and National Board of Medical Examiners, Phila., PA., described the use of a network to provide access to models that simulate medical behavior of patients. These models are used primarily for teaching and testing physicians. The network provides an interface by which varieties of terminals can connect to and access these models. Other data bases exist to which access through a network may be desirable; however, these data bases have a "polyglot" of organizations making it presently impossible to use foreign data bases.

Hector Maynez, National Library of Medicine, described the MEDLINE system. This has 1000 journals on-line to which access can be made via a network. This network, as the one above, provides the interface for access by various terminals. In this network are four or five computers with other applications such as CAI, clinical diagnosis, etc.

Roy Beveridge, MITRE, presented the requirements for the WWMCCS (World Wide Military Command and Control System) Network. This network will

contain 25 nodes and have a data exchange rate of the order of 10,000,000 characters per day. Three type of data were formulated - query data with response on the order of seconds, daily exchange for updates and reports, and other data for weekly, monthly or as required reports.

Erika Perez, MITRE, discussed data management for the WWMCCS Network. The two problems are determining the location of desired data, and providing the proper security and reliability for vital data. The location of data bases will be indicated in directories which may automatically determine which segment is applicable to a query. The directory will contain lists of data bases, files users and programs. The directory can be centralized (all at one location), distributed (split into pieces but where each piece resides at one location) partially replicated (split into pieces but in which certain parts may be replicated at different locations) and completely replicated (the complete directory at all locations).

The data management system will have to deal with possibly different hardware systems and even different local data managements systems. One solution is to have a standard data management and data description language for transmission of requests and data in the network.

The system will have to provide capabilities for file transfer, queries, remote batch, and for user communication via a mail box. The security of the data is maintained by checking user id, terminal authorization, process authorization and data authorization.

Bob Brown, General Motors Research Lab., described the network of computers at the General Motors Research Center. This network at present consists of an IBM 360/67, a 360/65, a 370/165, three 1800's and a Sigma 5. All of these are primarily for graphics use except the 67 and the 165. An example of how data passes through the network was given. The styling department develops a design on an 1800. Data on this design is sent to the 67 for stress and shape analysis and the results returned to the 1800. After a design is developed, it is sent to the 65-1800 combination for detailed analysis for production. Many of the computers are running GM's own operating systems, and the network control consists of macros added to these operating systems. Interfacing is done by providing specific conversion modules to be called when the specific conversion is required. The 67 will eventually be replaced by a hierarchical multiprocessor based on the CDC Star-100.

Phil Messing, MITRE, is setting up an experiment to test the practicability of interfacing a network standard data management language with local data management systems. In this experiment, a user will make a request in the network language, this request will be transmitted to a node, and translated to the language of this local node. At present, three local systems have been selected to be used - MADAM at MIT, LISTAR and Lincoln Labs., and NASIS at NASA/Ames.

It is not expected that the common data language will be able to handle all possible requests that may be made. The language should be able to handle the most common requests, otherwise, some means of interaction may be set up

in order to allow the transmission of more information to the target system than the common language may allow, or finally, a user can utilize the local target language.

At a later stage in the experiment, a user will input a query, the local host will determine where the query is to be sent, the transmission takes place, it is accepted by the target node, translated to the target node's local language and processed.

Ernie Forman, MITRE, is developing a special, simple data management system specifically for the purpose of measuring and testing organizational techniques for control, directories, and files. The question to be answered is whether each of these three functions should be centralized, or distributed, how, and where. The initial experimental arrangement is to have the control and directory centralized at the Rand node, and the files to be distributed at UCSB, Rand, and BBN. The files are each split vertically and distributed, this organization chosen to present the more difficult case.

Dick Watson, SRI, described some extensions of NIC (Network Information Center) that he would like to see, and that would involve network data management facilities. The first would be the ability to process text from one text processor by another. Second, it would eventually be desirable to distribute the NIC journals. A first stage of this would be to have several NLS (Network Library System) systems around the network, each with its own journal. The problems with this first stage would be in coordination of numbering and in organization of the directory. A second stage would be

one in which the journal might reside, in part, on other than NLS systems.

A third extension is to enable the NLS System to use the results of some other cataloging or citation and bibliographic referencing systems as input to the NLS catalogs. The fourth extension would be to enable other data management systems to generate data of more general type and be usable by the NLS.

Phase II

The second phase of the meeting was a working meeting to try and organize the committee and try and set up an active working interest group.

The following names presently form the committee. These are the people who have shown active interest, and are engaged in related activities:

Douglas B. McKay	IBM Research (Chairman)
Abhay Bhushan	MIT
Ernie Forman	MITRE
Dorothy Hopkin	University of Illinois
Phil Messing	MITRE
A. P. Mullery	IBM Research
Erika Perez	MITRE
A. Shoshani	SDC
S. Taylor	MITRE
Bob Thomas	BBN
Frank Ulner	NBS
Dick Watson	SRI
Dick Winter	CCA

It would be very useful in follow-on meetings to have a representative from the Form Machine group. Discussions on various uses of the Form Machine by a Network Data Management facility are bound to come up in later meetings.

A member of the form machine group would be an asset to the Data Management Committee.

Discussion on network data management covered many aspects of the problem with a general discussion on just what people want to be able to do with a network data facility.

The following list, gleamed from the discussion, represents the possible stages of development:

1. Transmission Facility - The Network Data Control Facility (DCF) is able to route requests for files to the proper node. The location and name must be specified.
2. Location Catalog - The DCF now has available to it a catalog which contains the locations of the data sets to be used in the network. Requests for files may be made by name only, the location being determined by the DCF.
3. Description Catalog - Descriptions, as well as data sets can be transmitted in the network. It is assumed these descriptions exist as files at local nodes. A target node can make use of the description to properly convert the data set to its own format.
4. Data Conversion Modules - Data descriptions are received by this module of the DCF. Based on the descriptions, conversion programs are called or

generated which will transform a file to the form required by the target node.

5. File Access Command Interface - This module is able to convert a request for a file from a network data language to the local language at which the file is located.

6. Data Access - This module, an extension of the network data language and the interface modules, allows access to pieces of data as specified in the data language, and generates the proper local access commands.

7. Data Management Interface - This is the final stage, at which general types of commands can be interfaced to local data managements systems, providing general interaction among different data management systems at different nodes.

It was generally agreed that the ability to access all data and different data bases is a goal which is worth achieving. There was discussion in what is the best way to achieve this goal, and the actual implementation techniques that could be used to achieve this. It was agreed that the data base interfacing problem should be studied in more detail and several people more willing to write reports on a representative problem when they have more results from their work.

There was also a discussion concerning the data language and whether it is

suitable or not. One fact should be made clear, the results of this committee should not fail or succeed on the outcome of the data language question. The initial proposal recommends the Datalanguage as a de facto standard that will be adopted in the network because of its support and availability. The group should be able to recommend changes when changes are shown to be necessary.

The Datalanguage discussion did point out the need for having data set descriptions cataloged and referrable by name - D. Winter, said that he would look into this problem.

The proposal (RFC 304) for a network data facility should be read again and discussed in more detail at our next meeting. The proposal says we can implement and achieve a stage 3 capability with what we know today. It would be a useful stepping stone to a stage 5 and stage 6 capability.

Related to the stages of development described above the following studies are now in progress and will help us answer pertinent questions.

A. Bhushan is studying a stage 1 type of network operation with extension in local catalogs to contain entries of network data sets of interest locally, to enable automatic calls to foreign data sets.

E. Perez will be studying the network catalog structure in more detail and will publish an RFC on her work.

Many questions were raised about the use of the data language as a network standard. There are two people that have volunteered writing up their investigations of this important study.

Frank Ulmer will be looking at various data management systems to see if their data structures are describable in terms of the Datalanguage. In addition, the NIC represents one important network data base that could be distributed through the network. Dick Watson will try to describe the NLS Journal structure in terms of the Datalanguage.

If there are any other people in the ARPA network or outside within hearing distance of this memo who may know about any real or potential applications of data sharing in a network, please submit an RFC in a letter to someone associated with the Data Management committee describing it.

Appendix

Meeting Attendees

William Benedict	USAFETAC Bldg. 159 Navy Yard Annex Wash. D.C.
Roy Beveridge	MITRE
Abhay Bhushan	MIT, Project Mac, Cambridge, Mass.
Bob Brown	General Motors Research Lab.
Elizabeth Fong	National Bureau of Standards, Wash. D.C.
Ernie Forman	MITRE
Glen Grazier	USAFETAC Bldg. 159 Navy Yard Annex Wash. D.C.
Dorothy Hopkin	U. of Ill., Adv. Comp. Bldg., Urbana, Ill.
Hector S. Maynez	National Library of Medicine
Doug B. McKay	IBM Research Center
Phil Messing	MITRE
Al Mullery	IBM Research Center
Erika Perez	MITRE
John Senior	Univ. of Penn. and National Board of Medical Examiners, Phila. PA.
Arie Shoshani	SDC, 2500 Colorado Ave., Santa Monica, Cal.
Martin Snyderman	Smithsonian Science Info. Exch., Wash. D.C.
Eric Swarthe	National Bureau of Standards, Wash. D.C.
Suzanne Taylor	MITRE
Bob Thomas	BBN
Frank Ulmer	National Bureau of Standards, Wash. D.C.
Dick Watson	SRI
Richard Winter	Computer Corporation of America